Error-Aware Spatio-Temporal Aggregation in the Model Web

Christoph Stasch, Edzer Pebesma, Benedikt Graeler and Lydia Gerharz

Abstract Spatio-temporal aggregation of observed or predicted values for environmental phenomena is needed for fusing sensor data or coupling sensors and environmental models. However, estimates from sensors or environmental models can never represent our world precisely and are subject to errors. Hence, there is uncertainty in the estimates that needs to be considered in environmental model workflows. This chapter presents an approach for an error-aware spatio-temporal aggregation in the Web, where probabilistic uncertainties are used within a Monte Carlo simulation. The approach is applied in a Web-based model chain that provides uncertain crop yield predictions on field parcel level that are aggregated to larger regions.

1 Introduction

The Model Web envisions discovery and access of environmental observations and models using the internet as mediating platform (Geller and Turner 2007; Nativi et al. 2012). Where environmental models, even those of same domains, currently exist in parallel and do not benefit from each other, the Model Web could ease the coupling of such models. To achieve this vision, the environmental

C. Stasch (✉) · E. Pebesma · B. Graeler · L. Gerharz
Institute for Geoinformatics, University of Münster, Münster, Germany
e-mail: staschc@uni-muenster.de

E. Pebesma
e-mail: edzer.pebesma@uni-muenster.de

B. Graeler
e-mail: ben.graeler@uni-muenster.de

L. Gerharz
e-mail: gerharz@uni-muenster.de

D. Vandenbroucke et al. (eds.), *Geographic Information Science at the Heart of Europe*, Lecture Notes in Geoinformation and Cartography,
DOI: 10.1007/978-3-319-00615-4_12,

observations and models should be exposed via publicly available standardized Web service interfaces such as those defined by the Open Geospatial Consortium (Maue et al. 2011). Spatio-temporal aggregation (Jeong et al. 2004; Vega Lopez et al. 2005; Stasch et al. 2012) is needed in the Model Web for two reasons: The spatio-temporal resolution of the sensor output might not match the resolution required by a model and, when chaining environmental models, the resolution of the output of one model might differ from the resolution required by another model.

However, environmental observations and models are subject to error due to the observation methods or due to simplified representation of real world phenomena by models. As a result, there is uncertainty in observations and model results that needs to be considered (Heuvelink 1998). The UncertWeb project aims to provide tools for managing and communicating uncertainties in the Model Web (Bastin et al. 2013). Such an uncertainty-enabled Model Web requires an error-aware spatio-temporal aggregation that explicitly considers uncertainties in input data and allows to propagate uncertainties to the aggregated estimates. As uncertainty can be reduced by aggregation in model workflows, an error-aware spatio-temporal aggregation also provides means to adjust the uncertainty, for example by averaging out some of the variability in the data.

The core contribution of this chapter is an approach for an error-aware spatio-temporal aggregation in the Model Web relying on open standards. A probabilistic approach is chosen for representing the uncertainties and a Monte Carlo simulation is used to propagate uncertainties in aggregation processes. To provide error-aware aggregation processes in the Model Web, a common Web service interface is defined and implemented in a Web-based model workflow for predicting land-use and crop yield response to climatic and economic change in England (Jones et al. 2012).

The remainder of the chapter is structured as follows: Sect. 2 provides an overview of error-aware spatio-temporal aggregation. Afterwards, the approach for Web-based error-aware aggregation is presented in Sect. 3. The application of the approach in a case study for aggregating yield predictions is described in Sect. 4, followed by the presentation of results in Sect. 5 and a discussion of the approach in Sect. 6. In the last section, conclusions are drawn and next steps are presented.

2 Error Aware Spatio-Temporal Aggregation

An aggregation process computes a single value, an aggregate, for a group of attribute values using an aggregation function. The values are grouped by partitioning predicates. Spatio-temporal aggregation groups spatio-temporal features by spatial and/or temporal predicates and applies aggregation functions to those features in order to change the spatio-temporal resolution of datasets (Jeong et al. 2004; Vega Lopez et al. 2005; Stasch et al. 2012). An example of a spatio-temporal aggregation process is shown in Fig. 1a. Temperature observations gathered hourly at monitoring stations are aggregated temporally to daily maxima

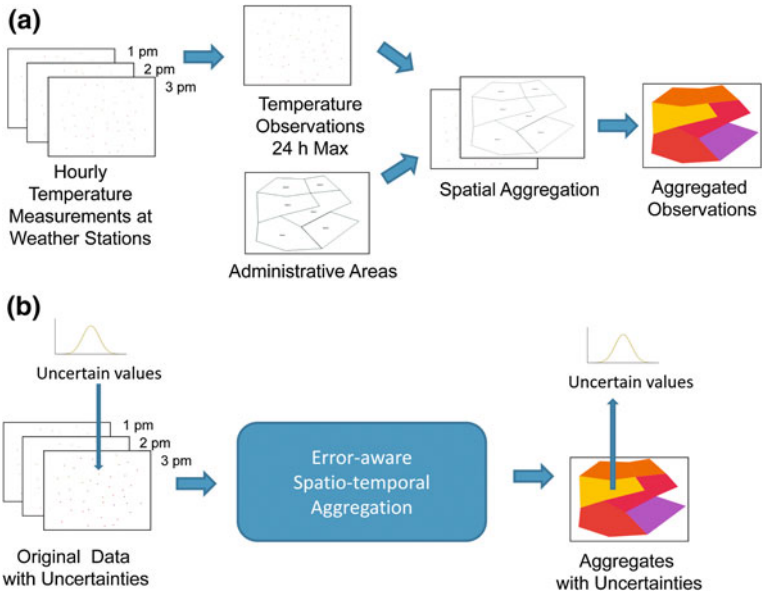

Fig. 1 Illustration of (error-aware) spatio-temporal aggregation (modified from Stasch et al. 2011). **a** Spatio-temporal aggregation. **b** Error-aware spatio-temporal aggregation

and spatially to means of spatial regions. The temporal grouping predicates consist of days, the temporal aggregation function is MAX, the spatial grouping predicates are spatial polygons (for instance, administrative boundaries) and the spatial aggregation function is MEAN. The grouping predicates as well as the aggregation functions might require additional specific input parameters to the aggregation process. For example, the spatial grouping predicates need to be defined by a set of polygons.

There are two possible sources of errors in spatio-temporal aggregation: (1) The input data of an aggregation process might be uncertain and cause errors in the aggregates or (2) the aggregation functions used for computing the values introduce uncertainties. An example of (2) is the aggregation of spatio-temporally distributed information on rainfall and soil moisture over a catchment during and after a rainfall event, which is aggregated to river discharge in order to predict floods. In this case, the aggregation function is a distributed hydrological model that can only approximate the true catchment response to a rainfall event, due to a simplified representation of the true aggregation process.

In this work, we focus on uncertainties in input data (1) that are aggregated using simple aggregation functions, e.g. mean, sum, or max, as illustrated in Fig. 1b. In case of spatio-temporal data, the uncertain input may be provided as a spatio-temporal random field $Y(q)$, where q is a spatio-temporal location. This random field is usually assumed to be normally distributed, i.e. $Y(q) \sim N(\mu(q), \Sigma)$, with $\mu(q)$ the mean vector for the locations and Σ the covariance matrix. In case an

aggregation function f is a non-linear function, e.g. computation of the maximum value, the expected value of the aggregates will typically differ from the aggregated expected values: $E[f(Y(q))] \neq f(E[Y(q)])$. Hence, aggregating the parameters of the input distributions in order to compute the probability distributions for the aggregates may introduce a bias. To avoid this, a Monte Carlo simulation approach is adopted to propagate the uncertainties in the aggregates (Heuvelink and Pebesma 1999). In case the inputs are provided as probability distribution functions (PDF) for each measurement value, realisations are generated from the input distributions and the aggregation process is run for each set of realisations resulting in a set of realisations for each output region that in turn approximates the target PDF.

The pseudocode for applying a Monte Carlo simulation is shown in Algorithm 1. The function y_i returns the i-th realisation value of a spatio-temporal random field $Y(q)$ at spatio-temporal location q within the target region R representing the grouping predicate. The realisation values per region are then used by the aggregation function f as inputs to compute the aggregate, for example, computing the sum. The actual aggregates for each spatio-temporal region R and i-th realisation r_i are returned by $\check{y}$. As an option, instead of returning all realisations of aggregates for each spatio-temporal region, summary statistics for the realisations, such as mean or the 95 %, may be computed by a function g as illustrated in Algorithm 2.

Algorithm 1 Spatio-temporal aggregation for multiple realisations

1: **for all** realisations r_i, i=1,...,n **do**
2: **for all** spatio-temporal regions R_j, j=1,...,m **do**
3: $\check{y}(R_j, r_i) = f(y_i(q_1), y_i(q_2), ..., y_i(q_p))$ *with* $\{q_1, ..., q_p\} \in R_j$
4: **end for**
5: **end for**

Algorithm 2 Aggregation with statistics computed from spatio-temporally aggregated realisations

1: **for all** spatio-temporal regions R_j, j=1,...,m **do**
2: $\bar{y}(R_j) = g(\check{y}(R_j, r_i), \check{y}(R_j, r_2), ..., \check{y}(R_j, r_n))$
3: **end for**

Besides allowing to propagate uncertainties with non-linear aggregation functions, the Monte Carlo simulation approach also allows for more flexibility than an analytical approach that is usually bound to a specific aggregation process (Heuvelink 1998). It also allows to consider input uncertainties for already existing deterministic aggregation processes without the need to change the underlying models of the aggregation processes. Spatio-temporal aggregation also provides a mean to control the uncertainty in model workflows: Given that there is variability in the data within the aggregation regions, aggregating the data to the mean of an

area, may reduce variability. However, the degree of variability reduction depends on the spatio-temporal autocorrelation (Gerharz and Pebesma 2012). The use case shown below illustrates that it also depends on the aggregation function used.

3 Error-Aware Aggregation in the Model Web

After introducing the general approach for an error-aware aggregation, the question remains how error-aware aggregation processes can be provided in the Model Web. Firstly, the input data needs to be provided with uncertainties and these need to be encoded in a standardized format (Sect. 3.1). Secondly, a common approach for providing and utilizing aggregation functionality in the Model Web needs to be defined that explicitly considers uncertainties (Sect. 3.2).

3.1 Formats for Spatio-Temporal Data with Uncertainties

In order to enable an error-aware spatio-temporal aggregation, the input data needs to contain uncertainty information. Up to now, if present at all, the uncertainty information is given in proprietary formats hindering a common approach and implementation of an error-aware aggregation. Hence, there is a need to provide common models and encodings for spatio-temporal data explicitly containing uncertainties. The Uncertainty Markup Language (UncertML) (Williams et al. 2009) has been developed as a common model and exchange format for probabilistic uncertainties. It allows to encode uncertainties as distributions, descriptive statistics or as a set of realisations. As UncertML does not explicitly define how to add spatial and temporal references to the uncertainties, there is a need for spatio-temporal models and exchange formats that support uncertainties.

To exchange uncertain spatio-temporal data in the Model Web, two common formats are defined. For vector data, the Uncertainty-enabled Observations & Measurements (U-O&M) format integrates UncertML with Observations & Measurements (O&M) (ISO 2010; Stasch et al. 2012), a common format for spatio-temporal observations and model results. Uncertainty can either be provided as additional metadata or as the result of an observation. U-O&M can be serialized in different formats such as XML, JSON, or plain text, such as comma separated values (csv), and hence be used for exchanging uncertain spatio-temporal data in the Model Web. The O&M format also allows to be used across different spatio-temporal aggregation levels of observations. In this work, we use observations with uncertain results as shown in Fig. 2 in XML format. The uncertainty is encoded as UncertML realisations in the result of the observation. While O&M is well suited for vector-based spatial data, NetCDF is a well-established format for gridded/raster data. NetCDF-U (Bigagli and Nativi 2011) has been defined to encode

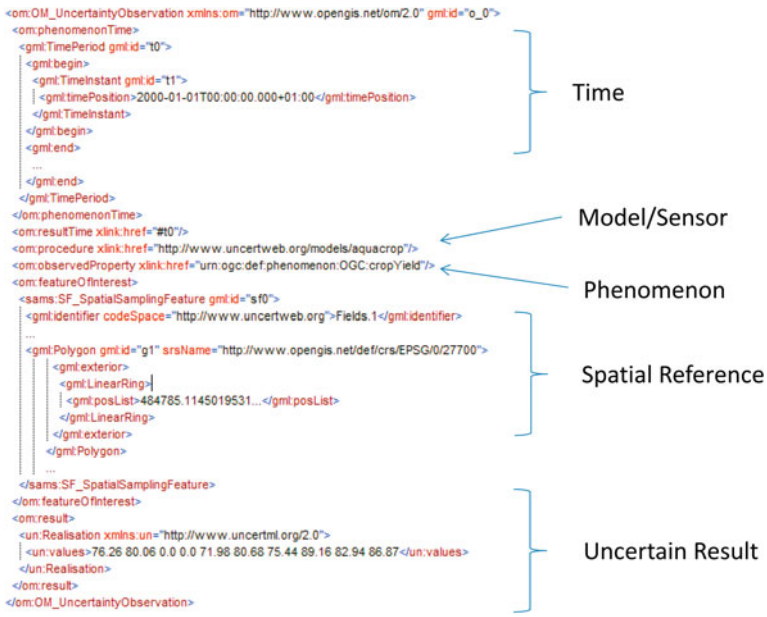

Fig. 2 Encoding of an observation that contains realisations of yield predictions for a field polygon in the year 2000

uncertainties in NetCDF (Domenico 2011) and is utilized in the Spatio-temporal Aggregation Service (STAS) for aggregating uncertain gridded data.

3.2 Error-Aware Spatio-Temporal Aggregation Service

To provide spatio-temporal aggregation functionality in the Model Web, we are extending the STAS that has been introduced by Stasch et. al. as an aggregation service for the Sensor Web (Stasch et al. 2012). As the Sensor Web envisions the tasking of sensors and the exchange of sensor data in the Web (Bröring et al. 2011), the Model Web may be seen as an extension that allows the discovery, access, and execution of environmental models and not just sensors. The overall concept of the STAS for the Model Web is illustrated in Fig. 3. The STAS is defined as a profile of the OGC Web Processing Service (WPS) (Schut 2007) and can be utilized as a mediator that transforms data from one resolution to another. The input data can be provided as output of model services, data sources, or as resources on a Web server. The STAS itself can then be invoked by end-users, model services, or orchestration engines and the aggregated data can in turn be directly published to model or data services or stored as a resource on a Web server.

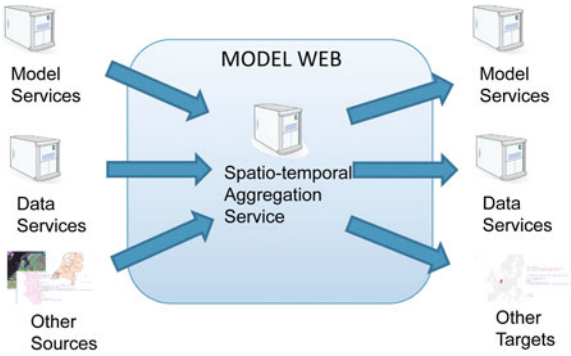

Fig. 3 Role of the spatio-temporal aggregation service in the model web. It acts as a mediator between data and model services in the model web, if an aggregation is required to fit outputs to other inputs. The aggregated data can be published via data services again

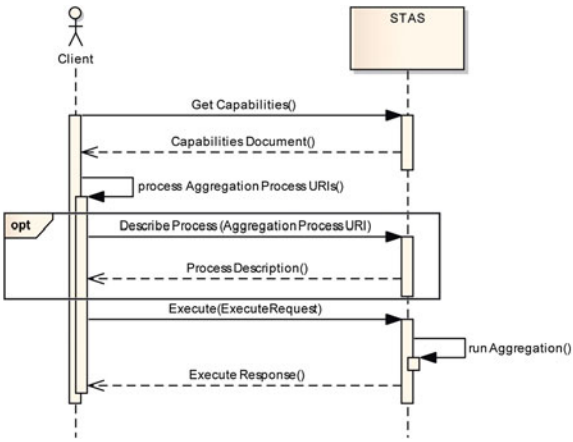

Fig. 4 UML sequence diagram showing the usual interaction pattern between a client and the STAS

The usual interaction pattern between clients and the aggregation service is shown as sequence diagram in the Unified Modeling Language (UML) in Fig. 4. The grouping predicates and aggregation functions of an aggregation process are identified in Unified Resource Identifiers (URI) of the aggregation processes. Therefore, we are utilizing the URI scheme as defined in (Stasch et al. 2012). All URIs of the available aggregation processes are listed in the service description (`Capabilities` document) that can be retrieved by the `GetCapabilities` operation. If a detailed description of a specific aggregation process including all input and output parameters is needed by the client, it can be retrieved using the `DescribeProcess` operation. To actually run an aggregation, the `Execute` operation needs to be invoked by passing an `ExecuteRequest` to the service. The request contains all necessary input parameters such as the input data (or a pointer to the data), parameters of the grouping predicates or of the aggregation functions. After aggregation, the `ExecuteResponse` can directly return the aggregated data in a requested format to the client or pass a reference, in case the aggregated data is inserted in another data service or stored on a server.

Table 1 Common input parameters of aggregation processes provided by the STAS for the model web

Input parameter Name	Cardinality	WPS input Type	Description
Identifier	1	URI	Identifier of the aggregation processes that should be run; defines the grouping predicates and aggregation functions
Variable	0...*	LiteralData	Name of variables (e.g. air temperature) that should be aggregated in case the input data contains several variables
InputData	1	ComplexData	Data that should be aggregated
SpatialFirst	0...1	Boolean	Indicates whether spatial aggregation should be done first (true) or not (false) in case of non-linear aggregation functions for space and/or time
TargetServer	0...1	LiteralData	Endpoint of the server, to which aggregated data should be written
TargetServerType	0...1	LiteralData	Type of server to which the aggregated data should be written

Common input parameters are defined for all aggregation processes in the Model Web as listed in Table 1. Depending on the grouping predicates and aggregation functions, additional parameters can be defined for particular aggregation processes. For example, the process introduced in Sect. 2 requires the additional parameter `FeatureCollection` that contains the polygons for the spatial grouping predicates and `duration` that defines the duration (24 h) for the temporal grouping predicates. The spatial and temporal references of the aggregates are defined by the parameters of the grouping predicates. The result of an aggregation execution not only provides the aggregated data, but also additional provenance information by pointing to an instance of a specific aggregation process description. This aggregation process description includes information about the predicates and aggregation functions used. In addition, the aggregated data points to the original data from which the aggregates have been computed.

A Monte Carlo simulation approach is used to propagate uncertainties in the STAS. The interaction pattern of an error-aware aggregation process in the STAS is shown in Fig. 5. Clients can indicate, whether a Monte Carlo simulation for the aggregation process should be run or not by passing the optional `NumberOf-Realisations` in an `Execute` request. If this parameter is present, the uncertain data has to be provided in the `InputData` parameter of the `Execute` request either as PDFs or as realisations. In case the uncertain inputs are PDFs, realisations are taken from the PDFs using the Uncertainty Transformation Service (UTS). The UTS is an external Web service for transforming uncertainties from one representation into another (Pross et al. 2012). For example, the UTS allows to convert from a normal distribution to a set of realisations. Then, for each Monte Carlo realisation, the aggregation is executed using an aggregation engine, e.g. the R software (R Development Core Team 2011), resulting in a set of samples of

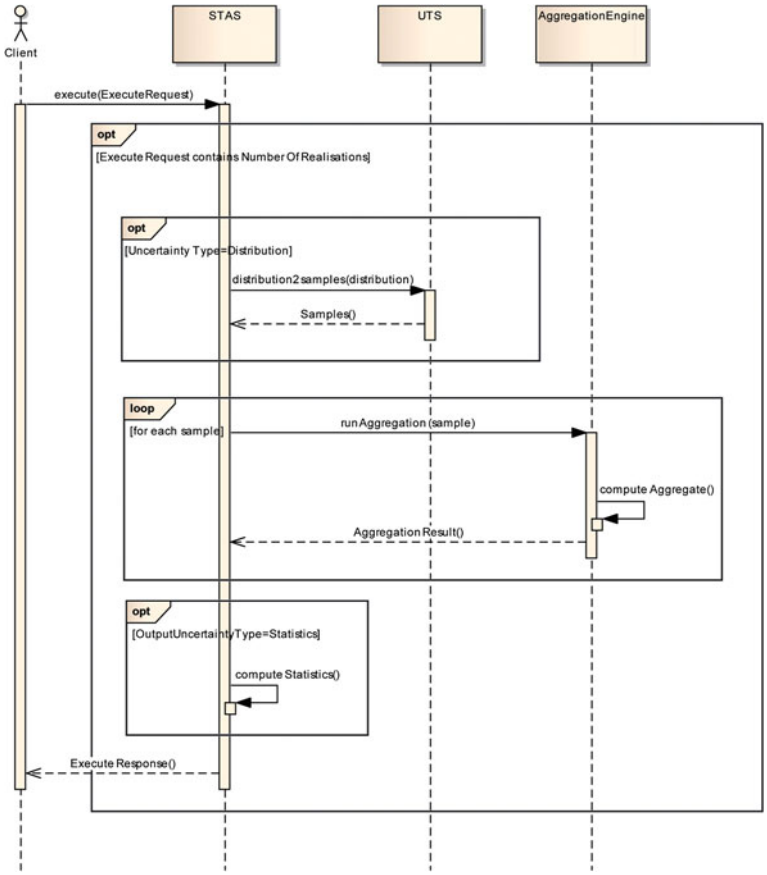

Fig. 5 UML sequence diagram showing the interaction pattern of an error-aware aggregation process between client, spatio-temporal aggregation service (STAS), uncertainty transformation service (UTS) and an aggregation engine

aggregated data. From the Web service, the aggregated data can either be retrieved as the full set of spatio-temporally aggregated realisations or as summary statistics of the realisations. Table 2 shows the additional input parameters for all error-aware aggregation processes. The `NumberOfRealisations` parameter is needed by all processes and defines the number of Monte Carlo simulation runs. In addition, the `OutputUncertaintyType` can define additional summary statistics that should be returned for the set of realisations of aggregated data. The `OutputUncertaintyType` parameter has to use the identifiers (URLs) of UncertML for the different statistics. The `InputData` parameter that is inherited from the common input paramters of the aggregation processes (Table 1) has the restriction to contain either realisations or distributions defined for UncertML in its inputs. In order to avoid errors, when clients are sending other uncertainty types or data without uncertainties to the service, the additional metadata element

Table 2 Additional parameters of error-aware aggregation processes

Input parameter name	Cardinality	WPS input type	Description
NumberOfRealisations	0..1	LiteralData	Number of Monte Carlo simulation runs
OutputUncertaintyType	0..*	LiteralData	The types of uncertainties as defined by UncertML in which the aggregated outputs should be provided. Per default, the aggregated data is provided as realisations, but also descriptive statistics of the realisations such as mean or standard deviation can be requested

`variable-uncertainty-types`, defined in the metadata conventions of the UncertWeb project, is nested in the `ows:Metadata` element of the `InputData` parameter in an aggregation process description is defined. The `variable-uncertainty-types` shall contain URLs of the UncertML dictionary for the supported uncertainty types. Besides the tag `variable-uncertainty-types`, several additional metadata tags, for example for the resolution of raster data, have been defined in the UncertWeb project and can also be used with other Web services than those defined by the OGC as described in Jones et al. (2012).

4 Case Study

This section describes a case study in which our approach is applied. The section starts with a description of the application scenario (Sect. 4.1) followed by a description of the Web service implementation (Sect. 4.2).

4.1 Application Scenario

The Food and Environment Research Agency[1] of the UK has established an environmental model workflow that is used to estimate land-use and crop yield responses to climatic and economic change. This model workflow has been extended to consider uncertainties and has been deployed via Web services in the internet (Jones et al. 2012). The model workflow estimates yields per field parcel for certain crop types, e.g. wheat or potatoes. The uncertainty in the yield predictions is propagated by running the yield model a number of times resulting in a number of yield realisations for each field per year.

[1] http://www.fera.defra.gov.uk/

Fig. 6 Excerpt from an overlay of fields and regions. Fields are shown in *white colour* and regions are shown in *grey colours*. The fields may be contained in several regions and some parts of the regions are not covered by fields

For privacy reasons and to provide an overview on a larger scale, the field estimates need to be aggregated to spatial regions. Thereby, the fields might be contained in several regions and some places in the regions might not be a field, e.g. urban areas or forests, as shown in Fig. 6.

Following our definition of spatio-temporal aggregation, the spatial grouping predicates are spatial regions and the spatial aggregation function f is defined as follows:

$$f(R) = \sum_{i=1}^{N_R} (x_i \times A_{iR}) \tag{1}$$

with R a spatial region over which we aggregate, N_R the number of field parcels intersecting R, x_i the estimated yield per hectar, and A_{iR} the spatial intersection area of field parcel i and region R. The yield prediction per hectar is multiplied with the area that intersects a region and for each region, the results are summed resulting in the total yield for each region. The data used in the case study consists of one thousand realisations of yield for 24 field parcels for the year 2012 that are aggregated to eight regions. For privacy reasons, not the real field parcel data are used, but artificially created parcels and regions were generated by a random process. As the data does only represent the year 2012, it does not need to be grouped temporally before executing the aggregation, though this is supported by the service implementation.

4.2 Web Service Implementation

The STAS is implemented as an extension of the 52° North WPS.[2] Therefore, an `AbstractAggregationProcess` class has been implemented that provides utility methods for accessing the common parameters of all aggregation processes. Several aggregation processes realizing the `AbstractAggregationProcesses` are implemented for vector and/or raster data. Three classes have been defined for the implementation of our error-aware aggregation approach as shown in Fig. 7. The `AbstractUncertainAggregationProcess` provides utility methods for the common inputs of error-aware aggregation processes and defines an additional abstract method `runMonteCarlo` that needs to be implemented by every subclass.[3]

To enable an aggregation as described in the previous Sect. 4.1, the class `Polygon2PolygonWeightedSum` has been implemented that extends the `AbstractUncertainAggregationProcess`. In addition, a `Polygon2PolygonMean` class is available to compute the arithmetic mean of the yields per region. For implementing the aggregation, a hash-based approach as described in Jeong et al. (2004) has been implemented in Java using the JTS library as follows: First, the input data is grouped by time and for each time, the input collection is stored in a hash map. Afterwards, the spatial grouping predicate (spatial intersection of the input features and the target regions) is checked. If there is an intersection, the realisations of intersecting features are cached with the intersection areas as weights for each target region using a hash map again. Then, the aggregation is executed for each target region several times until the number of realisations is reached and, depending on the requested uncertainty types, different statistics of the realisations per target region are computed.

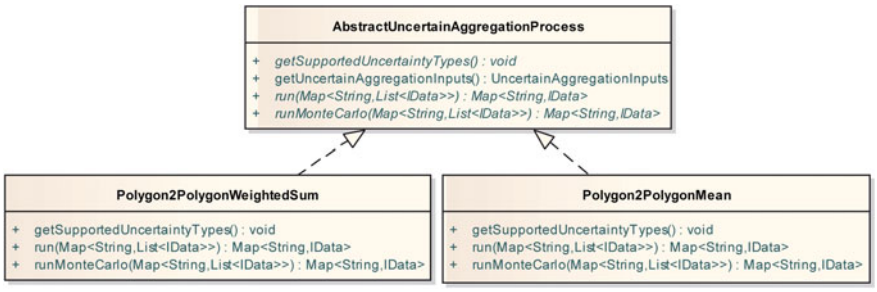

Fig. 7 UML class diagram of the two additional classes that are implemented for the error-aware aggregation service

[2] http://52north.org/communities/geoprocessing/wps/

[3] The classes are provided under the GNU General Public Licence (GPL) v2 licence as part of the STAS implementation at https://svn.52north.org/svn/geostatistics/main/uncertweb/stas/trunk

The uncertain yield predictions are provided in the U-O&M format with UncertML `ContinuousRealisations` as observation values for each field. The STAS runs the aggregation for each realisation of yield values per fields and hence produces 1,000 realisations for the aggregated yields per region. These are then returned again in the U-O&M format. The request/response encoding is automatically done by an additional component of the WPS framework developed in this work to support uncertain spatio-temporal inputs and outputs[4] (Sect. 3.1).

5 Results

Providing error-aware aggregation functionality in a standardized Web service allows for exchanging the aggregation methods in a flexible way in Web-based model workflows. In addition, the extension for Monte Carlo simulation allows for propagating the uncertainties during the aggregation.

Aggregation processes have been executed for computing the weighted sums and the mean of the yield values per region. Figure 8 shows the visualisation of the aggregated estimates in the UncertWeb visualisation client (Gerharz et al. 2012).

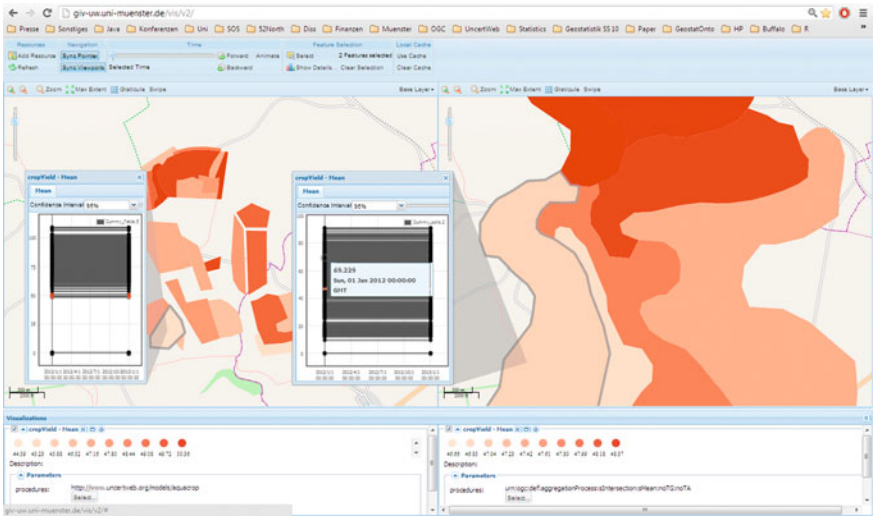

Fig. 8 Screenshot of the UncertWeb visualisation client visualising the non-aggregated (*left*) and aggregated yield estimates (*right*). Realisations of yield estimates can be visualised for specific fields and regions

[4] The input and output extension of the 52N WPS framework is acccessible as a separate package at https://svn.52north.org/svn/geostatistics/main/uncertweb/52n-wps-io-uncertweb/trunk and can also be used by other WPS implementations for uncertain data.

Table 3 Descriptive statistics of original data and aggregation results computed with mean as aggregation function

Description	Mean of realisation means	Standard deviations	Coefficient of variation
Non-aggregated yield estimates [in tonnes per hectar]	5.59	3.77	0.67
Aggregated yield estimates [in tonnes per hectar]	5.56	1.83	0.33

Table 4 Descriptive statistics of original data and aggregation results computed with weighted sum as aggregation function

Description	Mean of realisation means	Standard deviations	Coefficient of variation
Non-aggregated yield estimates [in tonnes]	44.04	29.69	0.67
Aggregated yield estimates [in tonnes]	155.54	49.94	0.32

For comparison, the non-aggregated as well as the aggregated estimates can be visualized. In this example, the realisation means of the fields as well as of the aggregates are shown in the map. For specific polygons all realisations are visualised in a popup. The information about the process that generated the data is given in the description of the layers. On the right side, for example, the URN shown in the procedures element is the identifier of the aggregation process.

Table 3 shows the descriptive statistics of the aggregation with mean as aggregation function. Firstly, means and variances of the yield realisations have been computed. Then, the means and variances have been derived from the realisation statistics. As expected, the mean yield per hectar is nearly the same for the fields as for the regions. The variability in the data is reduced by the mean aggregation, as the mean standard-deviation is reduced from 3.77 for the fields to 1.83 for the regions.

In order to compare the original field values with the weighted sum of the regions, the original values have been multiplied by the area of each field. While the aggregation to means of regions reduces the variability, the aggregation to weighted sums increases the variability as can be seen in Table 4. However, the coefficient of variation decreases in both cases.

6 Discussion

The approach of a Web based error-aware aggregation offers the following advantages: (1) a common way to communicate uncertain spatio-temporal data in the Web is defined, (2) the approach allows to change the resolution and

uncertainty in the data by aggregating them over space and time, (3) different error-aware aggregation processes can be accessed in a common way in the Web without the need to adopt workflow implementations for each aggregation process, (4) the standardized interface and the data formats allow for an integration in spatial information infrastructures, as they rely on standards used in these infrastructures.

The aggregation functionality provided by the aggregation service may be implemented, for example, in database systems (Jeong et al. 2004; Vega Lopez et al. 2005) or software systems such as R (Pebesma 2012) and then be provided by the aggregation service in the Web. While databases and other technologies usually only offer specific aspects needed in error-aware aggregation processes, the different technologies may be combined within the aggregation service. For instance, though there are databases for probabilistic data (Benjelloun et al. 2006), these databases do not provide spatio-temporal query functionality. An approach using Monte Carlo simulation for query evaluation is described in (Jampani et al. 2008), while they do not tackle the issue of aggregate queries on spatio-temporal entities. With our approach, the different technologies can be combined and provided in the Web via a common interface. In the case study, the input data has been transferred to the STAS. However, in case of big data, it may be more reasonable to tightly couple the aggregation functionality with the data sources as described for a coupling between the Sensor Observation Service and the STAS in our previous work (Stasch et al. 2012, p.117). The STAS interface can also be utilized in this case, but the parameter used for passing the inputs may then only identify data from the data source. This approach still allows to run the different aggregations in the Web.

Though the U-O&M format (Stasch et al. 2012) is defined for exchanging uncertain spatio-temporal vector data and NetCDF-U (Bigagli and Nativi 2011) is used to exchange uncertain spatio-temporal rater data, there is currently no generic standard for (uncertain) spatio-temporal data that is widely adopted. Hence, the (uncertain) spatio-temporal data needs to be converted, before it can be published in the Web and aggregated with the aggregation service. It needs to be explored, how well U-O&M and NetCDF-U map to other formats that are already in use and how uncertainty can be incoporated in such formats.

Another question is to which degree the process that generates the data relates to a sensor or to a model, or whether both concepts should be treated seperately. One would probably agree that a complex aggregation procedure such as a hydrological forecast model that aggregates measurements in space and time would not be considered as a sensor. However, observations such as discharge measurements usually have undergone a modelling procedure (Beven et al. 2012) and simple aggregations (and underlying models) are always part of technical sensors where the aggregation is done on a low abstraction level. Hence, there is still a need to clarify the semantics of the different concepts and, in a second step, to formalize them in order to be used in the Web for semantic interoperability (Sheth et al. 2008; Balazinska et al. 2007).

As our use case is based on realisations, the spatial autocorrelation does not need to be addressed in the Monte Carlo simulation. However, data formats defined in our previous work (Graeler and Stasch 2012), can be utilised to represent spatio-temporal random fields. Web services can then use the formats and provide spatio-temporal sampling procedures that consider the autocorrelation.

The current approach allows for the propagation of quantified uncertainties either provided as realisations or probability distributions. Further investigation is needed to check whether the current approach might be utilized in a scenario with categorical data. Furthermore, our approach requires that uncertainty information is available in the input data. Though there are already methods how to assess the uncertainty in measurements (Taylor 1997), assessing the uncertainty per observation or model output remains a challenging statistical and operational problem. In addition, most sensor data providers do not yet provide uncertainty information with the data. Hence, incentives have to be explored how to motivate data providers to make the uncertainties in their data explicit.

The implementation described in this chapter only includes aggregation processes, though the service interface of the STAS can also be used to provide disaggregation processes as described in (Bierkens et al. 2000). Finally, once there are more error-aware aggregation processes available in the Web that can be easily combined with model and sensor services, the question remains how to find those services and how to indicate that these aggregation services support uncertainty propagation. While we have introduced a description format for aggregation processes in our previous work, it has to be explored how these descriptions may be integrated in common approaches for sensor discovery (Jirka et al. 2009) and model web services (Nativi and Bigagli 2009).

7 Conclusion

This chapter presents an approach for an error-aware aggregation in the Model Web. For error propagation, Monte Carlo simulation is utilized. The uncertainty in the non-aggregated input data is provided as probability distributions, from which samples are taken, or is directly provided as samples. Then, for each sample an aggregation is carried out. Thus, the aggregation output consists of a set of realisations that in turn approximate the probability distribution of the aggregates. To deploy error-aware aggregation functionality in the Web, common data formats for uncertain spatio-temporal vector and raster data, U-O&M and NetCDF-U, are used and a Web service interface is defined as a profile of the OGC Web Processing Service. The application of the approach in a Web based model workflow for estimating crop yields in the UK shows that the approach allows for a flexible integration of aggregation processes and to propagate the uncertainties during aggregation. The approach also allows to tune the uncertainties in the data, depending on the aggregation function that is used.

Acknowledgments The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement no 248488. We are thankful to Jill Johnson and Sarah Knight from the Food and Environment Research Agency and Richard Jones from Aston University for the support during the integration of our approach in the yield prediction workflow.

References

Balazinska M, Deshpande A, Franklin M, Gibbons P, Gray J, Nath S, Hansen M, Liebhold M, Szalay A, Tao V (2007) Data management in the worldwide sensor web. Pervasive Comput IEEE 6(2):30–40 (April–June 2007)

Bastin L, Cornford D, Jones R, Heuvelink GBM, Stasch C, Pebesma E, Nativi S, Mazzetti P, Williams M (2012) Managing uncertainty in integrated environmental modelling frameworks: the uncertweb framework. Environ Model Softw 39:116–134

Benjelloun O, Sarma AD, Halevy A, Widom J (2006) ULDBs: databases with uncertainty and lineage. In: Proceedings of the 32nd international conference on very large data bases. VLDB '06, VLDB Endowment, pp 953–964

Beven K, Buytaert W, Smith LA (2012) On virtual observatories and modelled realities (or why discharge must be treated as a virtual variable). Hydrol Process 26(12):1905–1908

Bierkens M, Finke P, De Willingen P (2000) Upscaling and downscaling methods for environmental research. Kluwer Academic Publishers

Bigagli L, Nativi S, (eds) (2011) NetCDF uncertainty conventions (NetCDF-U). OGC 11–163. Open geospatial consortium, Inc, pp 17 (accessed 24 July 2012)

Bröring A, Echterhoff J, Jirka S, Simonis I, Everding T, Stasch C, Liang S, Lemmens R (2011) New generation sensor web enablement. Sensors 11(3):2652–2699

Domenico B (2011) OGC network common data form (NetCDF) core encoding standard version 1.0. OGC 10–090r3. Open geospatial consortium, Inc, pp 21 (Accessed on 01 Nov 2012)

Geller G, Turner, W.: The model web: a concept for ecological forecasting. In: Geoscience and Remote Sensing Symposium, (2007) IGARSS 2007. IEEE International. 2007:2469–2472

Gerharz L, Autermann C, Hopmann H, Stasch C, Pebesma E (2012) Uncertainty visualisation in the model web. European Geosciences Union (EGU) General Assembly

Gerharz L, Pebesma E (2012) Using geostatistical simulation to disaggregate air quality model results for individual exposure estimation on GPS tracks. Stoch Env Res Risk Assess 27:223–234

Graeler B, Stasch C (2012) Flexible representation of spatio-temporal random fields in the model web. European Geosciences Union (EGU) General Assembly

Heuvelink G (1998) Error propagation in environmental modelling with GIS. Taylor & Francis

Heuvelink G, Pebesma E (1999) Spatial aggregation and soil process modelling. Geoderma 89:47–65

ISO/TC211: ISO/FDIS 19156:2010: geographic information—observations and measurements. ISO/TC 211 (2010)

Jampani R, Xu F, Wu M, Perez LL, Jermaine C, Haas PJ (2008) MCDB: a Monte Carlo approach to managing uncertain data. In: Proceedings of the (2008) ACM SIGMOD international conference on Management of data. SIGMOD '08. New York, NY, USA, ACM, pp 687–700

Jeong SH, Fernandes AAA, Paton NW, Griffiths T (2004) A generic algorithmic framework for aggregation of spatio-temporal data. In: SSDBM '04: proceedings of the 16th international conference on scientific and statistical database management, Washington, DC, USA, IEEE Computer Society, p 245

Jirka S, Bröring A, Stasch C (2009) Discovery mechanisms for the sensor web. Sensors 9(4):2661–2681

Jones R, Cornford D, Bastin L (2012) UncertWeb processing service: making models easier to access on the web. Trans GIS 14(6):921–939

Maue P, Stasch C, Athanasopoulos G, Gerharz L (2011) Geospatial standards for web-enabled environmental models. Int J Spatial Data Infrastruct Res 6:145–167

Nativi S, Bigagli L (2009) Discovery, mediation, and access services for earth observation data. IEEE J Sel Top Appl Earth Observ Rem Sens 2(4):233–240

Nativi S, Mazzetti P, Geller GN (2012) Environmental model access and interoperability: the GEO model web initiative. Environ Model Softw 39:214–228. doi:10.1016/j.envsoft.2012.03.007

Pebesma E (2012) Spacetime: spatio-temporal data in R. J Stat Softw 51(7):1–30

Pross B, Gerharz L, Stasch C, Pebesma E (2012) Tools for uncertainty propagation in the model web using Monte Carlo simulation. In: Seppelt R, Voinov A, Lange S, Bankamp D (eds) Proceedings of the iEMSs sixth Biennial meeting: Managing resources of a limited planet. International congress on environmental modelling and software (iEMS 2012), international environmental modelling and software society (iEMSs)

R Development Core Team: R (2011) A Language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0

Schut P (2007) OpenGIS web processing service. OGC 05–007r7. Open Geospatial Consortium, Inc., 87pp. (Accessed on 24 July 2012)

Sheth A, Henson C, Sahoo S (2008) Semantic sensor web. IEEE Int Comput, pp 78–83

Stasch C, Foerster T, Autermann C, Pebesma E (2012) Spatio-temporal aggregation of European air quality observations in the sensor web. Comput Geosci 47:111–118

Stasch C, Autermann C, Foerster T, Pebesma E (2011) Towards a spatiotemporal aggregation service in the sensor web. Poster presentation. In: The 14th AGILE international conference on geographic information, science

Stasch C, Jones R, Cornford D, Kiesow M, Williams M, Pebesma E (2012) Representing Uncertainties in the Sensor Web. In: Proceedings of Workshop Sensing A Changing World

Taylor JR (1997) An introduction to error analysis: the study of uncertainties in physical measurements. University Science Books

Vega Lopez IF, Snodgrass RT, Moon B (2005) Spatiotemporal aggregate computation: a survey. IEEE Trans Knowl Data. Engineering 17(2):271–286

Williams M, Conford D, Bastin L, Pebesma E (2009) Uncertainty markup language (UncertML) (OGC 08–122r2)